CLAIMS

What is claimed is:

1. A method for generating a text analysis program for recognizing patterns appearing in text and extracting information from said patterns, the method comprising the steps of:

(a) providing a sample hierarchy, said sample hierarchy comprising samples of text;

(b) extracting at least one rule from said sample hierarchy, said rule describing how to process a portion of text;

(c) generating a pass from said rule, said pass containing instructions to operate a text analyzer; and

(d) constructing a text analyzer containing said pass.

2. The method of claim 1, wherein said rule is generalized into multiple rules and multiple passes.

3. The method of Claim 1, wherein multiple passes are added to said text analyzer.

4. The method of Claim 3, wherein said multiple passes are arranged in a cascading manner having a sequence of passes such that rules associated with a pass are applied to subsequent passes.

5. The method of Claim 1, wherein the samples are associated with offset values, said offset values identifying locations in a parse tree data structure, said parse tree containing concepts stored at locations identified by said offsets.

6. The method of Claim 4, further comprising the step of allowing a user to control the extraction of rules from the sample hierarchy

7. The method of Claim 5, further comprising the step of allowing a user to designate properties

associated with said rules, said properties controlling rule generation for a portion of the sample hierarchy.

8. The method of Claim 5, wherein said concepts are retrieved from said parse tree and processed to form said rule.

9. The method of Claim 6, further comprising the step of allowing a user to designate attributes associated with said rules, said attributes guiding the application of said rules.

10. The method of Claim 1, wherein multiple rules are generalized and merged into a single rule if there is a difference between the multiple rules.

11. The method of Claim 10, wherein said samples may be contained in a sample file.

12. A sample hierarchy data structure for use in a text analyzer system, said sample hierarchy comprising an index for storing samples, said samples comprising portions of text, said samples used to generate rules for identifying patterns appearing in text, said samples used to derive information from said identified patterns, said rules generated by parsing said text samples, said index organized such that passes comprising operational steps and rules are generated in an order wherein simple patterns are recognized by said text analyzer, and said recognized simple patterns are used by said text analyzer system and used to iteratively recognize more complex patterns.

13. A computer readable medium containing instructions which, when executed by a computer, generate a text analysis program for recognizing patterns appearing in text and extracting information from said patterns, by:

(a) providing a sample hierarchy, said sample hierarchy comprising samples of text;

(b) extracting at least one rule from said sample hierarchy, said rule describing how to process a portion of text;

(c) generating a pass from said rule, said pass containing instructions to operate a text analyzer; and

(d) constructing a text analyzer containing said pass.